The Positive Power of Prejudice: A Computational Model for MAS

Alessandro Sapienza, Rino Falcone and Cristiano Castelfranchi
Institute of Cognitive Science and Technologies, ISTC-CNR, Rome, Italy
{alessandro.sapienza, rino.falcone, cristiano.castelfranchi}@istc.cnr.it

***Abstract*— In MAS studies on Trust building and dynamics the role of direct/personal experience and of recommendations and reputation is proportionally overrated; while the importance of inferential processes in deriving the evaluation of trustees' trustworthiness is underestimated and not enough exploited.**

In this paper we focus on the importance of generalized knowledge: agents' categories. The cognitive advantage of generalized knowledge can be synthesized in this claim: "It allows us to know a lot about something/somebody we do not directly know". At a social level this means that I can know a lot of things on people that I never met; it is social "prejudice" with its good side and fundamental contribution to social exchange. In this study we experimentally inquire the role played by categories' reputation with respect to the reputation and opinion on single agents: when it is better to rely on the first ones and when are more reliable the second ones. Our claim is that: the larger the population and the ignorance about the trustworthiness of each individual (as it happens in an open world) the more precious the role of trust in categories. In particular, we want investigate how the parameters defining the specific environment (number of agents, their interactions, transfer of reputation, and so on) determine the use of categories' reputation.

This powerful inferential device has to be strongly present in WEB societies.

I. INTRODUCTION

In MultiAgent Systems (MAS) and Online Social Networks (OSN) studies on Trust building and dynamics the role of direct/personal experience and of recommendations and reputation (although important) is proportionally overrated; while the importance of inferential processes in deriving the evaluation of trustee's trustworthiness is underestimated and not sufficiently exploited (a part from the so called "transitivity", which is also, very often, wrongly founded).

In particular, generalization and instantiation from classes, categories [8] and analogical reasoning (from task to task and from agent to agent) really should receive much more attention. In this paper we focus on the importance of generalized knowledge: agents' categories. The cognitive advantage of generalized knowledge (building *classes*, *prototypes*, *categories*, etc.), can be synthesized in this obvious claim: "It allows us to know a lot about something/somebody we do not directly know" (for example, I never saw Mary's dog, but - since it is *a dog* - I know hundreds of things about it). At a social level this means that I can know *a lot of things on people that I never met*; it is social "prejudice" with its good side and fundamental contribution to social exchange. How can I trust (for drugs prescription) a medical doctor that I never met before and nobody of my friends knows? Because he is a doctor!

Of course we are underlining the positive aspects of generalized knowledge, its essential role for having information on people never met before and about whom no one gave testimony. The more rich and accurate this knowledge is, the more it is useful. It offers huge opportunity both for realizing productive cooperation and for avoiding risky interactions. The problem is when the *uncertainty about the features* of the categories is too large or it is too wide the *variability of the performers* within them. In our culture we attribute a negative sense to the concept of prejudice, and this because we want to underline how generalized knowledge can produce unjust judgments against individuals (or groups) when superficially applied (or worst, on the basis of precise discriminatory intents). Here we want rather to point out the positive aspects of the prejudice concept.

In this study we intend to explain and experimentally show the advantage of trust evaluation based on classes' reputation with respect to the reputation and opinion on single potential agents (partners). In an open world or in a broad population how can we have sufficient direct or reported experience on everybody? The quantity of potential agents in that population or net that might be excellent partners but that nobody knows enough can be high.

Our claim is that: the larger the population and the ignorance about the trustworthiness of each individual the more precious the role of trust in categories. If I know (through signals, marks, declaration, ...) the class of a given guy/agent I can have a reliable opinion of its trustworthiness derived from its class-membership.

It is clear that the advantages of such cognitive power provided by categories and prejudices does not only depend on recommendation and reputation about categories. We can personally build, by generalization, our evaluation of a category from our direct experience with its members (this happens in our experiments for the agents that later have to propagate their recommendation about). However, in this simulation we have in the trustor (which has to decide whom rely on) only a prejudice based on recommendations about that

category and not its personal experience.
After a certain degree of direct experiences and circulation of recommendations, the performance of the evaluation based on classes will be better; and in certain cases there will be no alternative at all: we do not have any evaluation on that individual, a part from its category; either we work on inferential instantiation of trustworthiness or we loose a lot of potential partners. This powerful inferential device has to be strongly present in WEB societies supported by MAS. We simplify here the problem of the generalization process, of how to form judgement about groups, classes, etc. by putting aside for example inference from other classes (higher or sub); we build opinion (and then its transmission) about classes on the bases of experience with a number of subjects of a given class.
First of all, we want to clarify that here we are not interested in steretypes, but in categories. We define steretypes as the set of features that, in a given culture/opinion, characterize and distinguish that specific group of people.
Knowing the stereotype of an agent could be expensive and time consuming. Here we are just interested in the fact that an agent belongs to a category: it has not to be a costly process and the recognition must be well discriminative and not-cheating. There should be visible and reliable "signals" of that membership. In fact, the usefulness of categories, groups, roles, etc. makes fundamental the role of the *signs* for recognizing or inferring the category of a given agent. That's why in social life are so important coats, uniforms, titles, badges, diplomas, etc. and it is crucial their exhibition and the assurance of their authenticity (and, on the other side, the ability to falsify and deceive). In this preliminary model and simulation let us put aside this crucial issue of indirect competence and reliability *signaling*; let us assume that the membership to a given class or category is true and transparent: the category of a given agent is public, common knowledge.
Differently from [2][10][17] in this work we do not address the problem of learning categorical knowledge and we assum that the categorizzation process is objective.
Similarly to [3], we give agents the possibility to recommend categories and this is the key point of this paper.
In the majority of the cases available in the literature, the concept of recommendation is used concerning recommender systems [1]. These ones can be realized using both past experience (content-based RS)[13] or collaborative filtering, in which the contribute of single agents/users is used to provide group recommendations to other agents/users.
Focusing on collaborative filtering, the concepts of similarity and trust are often exploited (together or separately) to determine which contributes are more important in the aggregation phase [14][18]For instance, in [7] authors provide a system able to recommend to users group that they could join in Online Social Network. Here it is introduced the concepts of *compactness* of a social group, defined as the weighted mean of the two dimensions of similarity and trust.
Even in [11] authors present a clustering-based recommender system that exploits both similarity and trust, generating two different cluster views and combining them to obtain better results.
Another example is [6] where authors use information regarding social friendships in order to provide users with more accurate suggestions and rankings on items of their interest.
A classical decentralized approach is referral systems [20], where agents adaptively give referrals to one another.
Information sources come into play in FIRE [12], a trust and reputation model that use them to produce a comprehensive assessment of an agent's likely performance. Here authors take into account open MAS, where agents continuously enter and leave the system. Specifically, FIRE exploits interaction trust, role-based trust, witness reputation, and certified reputation to provide trust metrics.
The described solutions are quite similar to our work, although we contextualized this problem to information sources. However we do not investigate recommendations with just the aim of suggesting a particular trustee, but also for inquiring categories' recommendations.

II. Recommendation and Reputation: Definitions

Let us consider a set of agents Ag_1, ..., Ag_n in a given world (for example a social network). We consider that each agent in this world could have trust relationships with anyone else. On the basis of these interactions the agents can evaluate the trust degree of their partners, so building their judgments about the trustworthiness of the agents with whom they interacted in the past.
The possibility to access to these judgements, through recommendations, is one of the main sources for trusting agents outside the circle of closer friends. Exactly for this reason recommendation and reputation are the more studied and diffused tools in the trust domain [15].
We introduce

$$\mathrm{Rec}_{x,y,z}(\tau) \tag{1}$$

where $x,y,z \in \{Ag_1, Ag_2, ..., Ag_n\}$, we call D the specific set of agents: $D \equiv \{Ag_1, Ag_2, ..., Ag_n\}$

and $0 \leq \mathrm{Rec}_{x,y,z}(\tau) \leq 1$

τ, as established in the trust model of [4], is the task on which the recommender *x* expresses the evaluation about *y*.
In words: $\mathrm{Rec}_{x,y,z}(\tau)$ is the value of *x*'s recommendation about *y* performing the task τ, where *z* is the agent receiving this recommendation. In this paper, for sake of simplicity, we do not introduce any correlation/influence between the value of the recommendations and the kind of the agent receiving it: the value of the recommendation does not depend from the agent to whom it is communicated.
So (1) represents the basic expression for recommendation.
We can also define a more complex expression of recommendation, a sort of *average recommendation*:

$$\sum_{x=Ag_1}^{Ag_n} \mathrm{Rec}_{x,y,z}(\tau)/n \tag{2}$$

in which all the agents in the defined set of agents express their individual recommendation on the agent *y* with respect the task τ and the total value is divided by the number of agents.

We consider the expression (2) as the <u>reputation</u> of the agent y with respect to the task τ in the set D.

Of course the reputation concept is more complex than the simplified version here introduced [5][16].

It is in fact the value that would emerge in the case in which we receive from each agent in the world its recommendation about y (considering each agent as equally reliable).

In the case in which an agent has to be recommended not only on one task but on a set of tasks (τ_1 , ..., τ_k), we could define instead of (1) and (2) the following expressions:

$$\sum_{i=1}^{k} \mathrm{Rec}_{x,y,z}(\tau_i)/k \qquad (3)$$

that represents the *x*'s recommendation about *y* performing the set of tasks (τ_1,..., τ_k), where *z* is the agent receiving this recommendation.

Imagine having to assign a meta-task (composed of a set of tasks) to just one of several agents. In this case the information given from the formula (3) could be useful for selecting (given the *x*'s point of view) on average (with respect to the tasks) the more performative agent *y*.

$$\sum_{x=Ag_1}^{Agn} \sum_{i=1}^{k} \mathrm{Rec}_{x,y,z}(\tau_i)/nk \qquad (4)$$

that represents a sort *of average recommendation* from the set of agents in D, about *y* performing the set of tasks (τ_1 , ..., τ_k).

We consider the expression (4) as the <u>reputation</u> of the agent y with respect the set of tasks (τ_1 , ..., τ_k), in the set D.

Having to assign the meta-task proposed above, the information given from the formula (4) could be useful for selecting on average (with respect to both the tasks and the agents) the more performative agent *y*.

A. Using Categories

As described above, an interesting approach for evaluating agents is to classify them in specific categories already pre-judged/rated and as a consequence to do inherit to the agents the properties of their own categories.

So we can introduce also the *recommendations about categories*, not just about agents (we discuss elsewhere how these recommendations are formed). In this sense we define:

$$\mathrm{Rec}_{x,Cy,z}(\tau) \qquad (5)$$

where $x \in \{Ag_1, Ag_2, ..., Ag_n\}$ as usual, and we characterize the categories $\{C_1,....,C_l\}$ through a set of features $\{f_{y1},..., f_{ym}\}$:

$\forall y \in \{Ag_1,...,Ag_n\} \exists c_y \in \{C_1,...,C_l\} \mid (C_y \equiv \{f_{y1},..., f_{ym}\}) \wedge (\{f_{y1},..., f_{ym}\} \in y)$

it is clear that there is a relationship between task τ, and the features $\{f_{y1},..., f_{ym}\}$ of the C_y category. In words we can say that each agent in *D* is classified in one of the categories $\{C_1,....,C_l\}$ that are characterized from a set of features $\{f_1,..., f_m\}$; as a consequence each agent belonging to a category owns the features of that category.

$0 \le \mathrm{Rec}_{x,Cy,z}(\tau) \le 1$

In words: $\mathrm{Rec}_{x,Cy,z}(\tau)$ *is the value of x's recommendation about the agents included in category C_y when they perform the task τ*, (as usual *z* is the agent receiving this recommendation).

We again define a more complex expression of recommendation, a sort of *average recommendation*:

$$\sum_{x=Ag_1}^{Ag_n} \mathrm{Rec}_{x,Cy,z}(\tau)/n \qquad (6)$$

in which all the agents in the domain express their individual recommendation on the category C_y with respect the task τ and the total value is divided by the number of the recommenders.

We consider the expression (6) as the <u>reputation</u> of the category C_y with respect the task τ in the set D.

Now we extend to the categories, in particular to C_y, the recommendations on a set of tasks (τ_1, ..., τ_k):

$$\sum_{i=1}^{k} \mathrm{Rec}_{x,Cy,z}(\tau_i)/k \qquad (7)$$

that represents *the recommendation value of the x's agent about the agents belonging to the category C_y when they perform the set of tasks (τ_1,..., τ_k).*

Finally, we define:

$$\sum_{x=Ag_1}^{Agn} \sum_{i=1}^{k} \mathrm{Rec}_{x,Cy,z}(\tau_i)/nk \qquad (8)$$

that represents *the value of the reputation of the category C_y (of all the agents y included in C_y) with respect the set of tasks (τ_1,..., τ_k), in the set D.*

B. Definition of Interest for this Work

In this paper we are in particular interested in the case in which *z* (a new agent introduced in the world) asks for recommendation to *x* ($x \in D$) about an agent belonging to its domain D_x for performing the task τ (D_x is a subset of *D*, it is composed by the agents that *x* knows). *x* will select the best evaluated *y*, with $y \in D_x$ on the basis of formula:

$$\max_{y \in D_x}(\mathrm{Rec}_{x,y,z}(\tau)) \qquad (9)$$

where $D_x \equiv \{Ag_1, Ag_2, ..., Ag_m\}$, D_x includes all the agents evaluated by *x*. They are a subset of *D*: $D_x \subseteq D$.

In general *D* and D_x are different because *x* does not necessarily know (has interacted with) all the agents in *D*.

z asks for recommendations not only to one agent, but to a set of different agents: $x \in D_z$ (D_z is a subset of *D*, to which z asks for reputation), and selects the best one on the basis of the value given from the formula:

$$\max_{x \in D_z}(\max_{y \in D_x}(\mathrm{Rec}_{x,y,z}(\tau))) \qquad (10)$$

$D_z \subseteq D$, z could ask to all the agents in the world or to a defined subset of it (see later).

We are also interested to the case in which *z* ask for recommendations to *x* about a specific *agents' category* for performing the task τ. *x* has to select the best evaluated C_y among the different $C_y \in \{C_1,....,C_l\}$ x has interacted with (we are supposing that each agent in the world *D*, belongs to a category in the set $\{C_1,....,C_l\}$).

In this case we have the following formulas:

$$\max_{Cy \in D_x}(\mathrm{Re}c_{x,Cy,z}(t)) \qquad (11)$$

that returns the category best evaluated from the point of view of an agent (x). And

$$\max_{x \in D_z}(\max_{Cy \in D_x}(\mathrm{Re}c_{x,Cy,z}(t))) \qquad (12)$$

that returns the category best evaluated from the point of view of all the agents included in D_z.

III. Computational Model

A. General Setup

In order to realize our simulations, we exploited the software NetLogo [19].

In every scenario there are four general categories, called Cat1, Cat2, Cat3 and Cat4, composed by 100 agents per category.

Each category is characterized by:

1. an **average value of trustworthiness**, in range [0,100];
2. an **uncertainty value**, in range [0,100]; this value represents the interval of trustworthiness in which the agents can be considered as belonging to that category.

These two values are exploited to generate the **objective trustworthiness** of each agent, defined as *the probability that, concerning a specific kind of required information, the agent will communicate the right information*.

Of course the trustworthiness of categories and agents is strongly related to the kind of requested information/task. Nevertheless, for the purpose of our it is enough to use just one kind of information (defined by τ) in the simulations. The categories' trustworthiness of Cat1, Cat2, Cat3 and Cat4 are fixed respectively to 80, 60, 40 and 20% for τ. What changes through scenarios is the uncertainty value of the categories: 1, 20, 50, and 80%.

B. How the simulations work

Simulations are mainly composed by two main steps that are repeated continuously. In the first step, called **exploration phase**, agents without any knowledge about the world start experiencing other agents, asking to a random 3% of the population for the information P. Then they memorize the performance of each queried agent both as individual element and as a member of its own category.

The performance of a agent can assume just the two values 1 or 0, with 1 meaning that the agent is supporting the information P and 0 meaning that it is opposing to P. For sake of simplicity, we assume that P is always true.

The exploration phase has a variable duration, going from 100 ticks to 1 tick. Depending on this value, agents will have a better or worse knowledge of the other agents.

Then, in a second step (**querying phase**) we introduce in the world a trustor (a new agent with no knowlegde about the trustworthiness of other agents and categories, and that has the necessity to trust someone reliable for a given informative task: in our case τ). It will select a given subset of the population, going from 100% to 5%, and it will query them. In particular, the trustor will ask them for the best category and the best trustee they have experienced.

In this way, the trustor is able to collect information about both the best recommended category and agent.

It is worth noting that the trustor collects information from the agents considering them equally trustworthy with respect to the task of "providing recommendations". Otherwise it should weigh differently these recommendations. In practice the agents are sincere.

Then it will select a randomly chosen agent belonging to the best recommended category and it will compare it, in terms of objective trustworthiness, with the best recommended individual agent (trustee).

The possible **outcomes** are:

- **trustee wins (t_win)**: the trustee selected with individual recommendation is better than the one selected by the means of category; then this method gets one point;
- **category wins (c_win)**: the trustee selected by the means of category is better than the one selected with individual recommendation; then this method gets one point;
- **equivalent result**: if the difference between the two trustworthiness values is not enough (it is under a threshold), we consider it as indistinguishable result. In particular, we considered the threshold of 3% as, on the basis of previous test simulations, it has resulted a resonable value.

These two phases are repeated 500 times for each setting.

IV. Simulations Results

In these simulations we present a series of scenarios with different settings to show when it is more convenient to exploit recommendations about categories rather than recommendations about individuals, and vice versa.

We also present the "all-in-one" scenario, whose peculiarity is that the exploration lasts just 1 tick and in that tick every agent experiences all the others. Although this is a limit case, very unlikely in the real world, it is really interesting as each agent has not a good knowledge of the other agent as individual elements (it experienced them just one time), but it is able to get a really good knowledge of their categories, as it has experienced them as many times as the number of agents for each category. This is an explicit case in which agents' recommendations about categories are surely more informative than the ones about individuals.

In particular, we will represent this value:

$$\frac{c\_win}{c\_win + t\_win} \qquad (13)$$

In words, this ratio shows how much categories' recommendation is useful if compared to individual recommendation.

Simulations' results are presented in a graphical way, exploiting 3D shapes to represent all the outcomes. These shapes are divided into two area and represented with two different colors:

- the part over 0.5, in which prevails the category recommendation;
- the one below 0.5, in which prevails the individual recommendation.

These graphs represent an useful view about the utility of the categorial role in the different interactional and social contexts.

For each value of uncertainty, we explored 40 different settings, considering all the possible couple of **exploration phase** and **queried trustee percentage**, where:

- exploration phase ∈ {all-in,1,3,5,10,25,50,100};
- queried trustee percentage ∈ {5,10,25,50,100}.

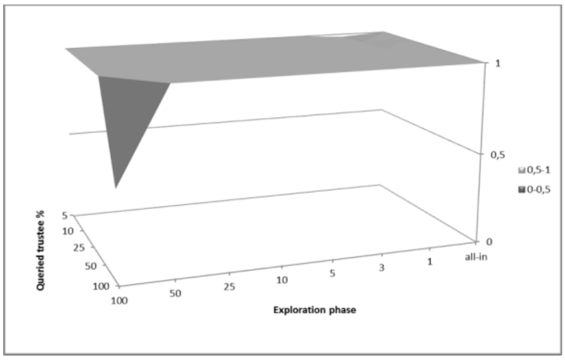

Figure 1. Outcomes for 1% of categories' uncertainty

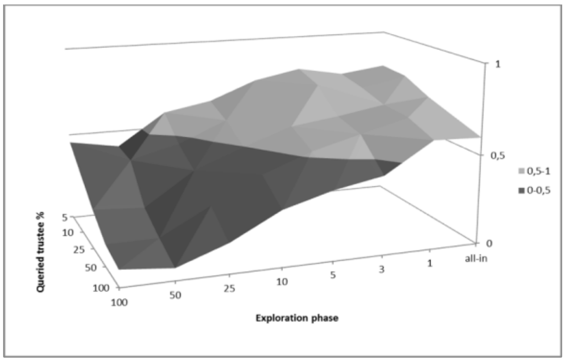

Figure 2. Outcomes for: 20% of categories' uncertainty

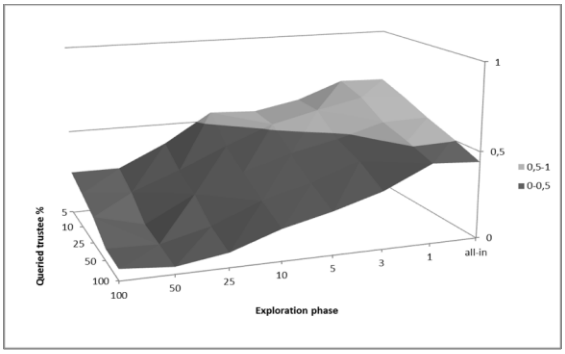

Figure 3. Outcomes for: 50% of categories' uncertainty

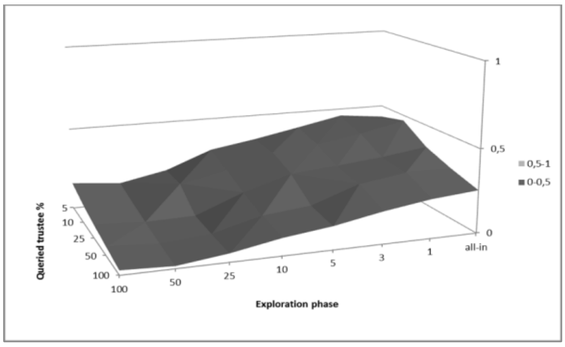

Figure 4. Outcomes for 80% of categories' uncertainty

The part in which category recommendation wins over individual recommendation is represented in light grey. Conversely, the part in which individual recommendation wins is represented in dark grey.

Through these graphs we identify three effects that influence the outcome. The **first effect** is due to categories' uncertainty: the less it is, the more is the utility of using categories; the more it is, the less categories will be useful. It is not possible to notice this effect just looking at one picture. On the contrary, looking at the overal picture one can notice that the curves of the graphs lower, going from a maximal value in **Figure 1** to a minimal value in **Figure 4**.

The **second effect** is due to exploration phase. The longer it is the more individual recommendations are useful; the less it lasts the more category recommendations are useful.

The **third effect** is introduced by the queried trustee percentage, that acts exactly as the exploration phase: the higher the percentage of queried agents, the more individual's recommendations are useful; the less it is, the more categories' recommendations are useful.

The exploration phase's length and the queried agents' percentage occur in all the four graphs and cooperate in determining respectevely the degree of knowledge (or ignorance) in the world and the level of inquire about this knowledge. In particular, with "the knowledge in the world" we intend how the agents can witness the trustworthiness of the other agents or their aggregate, given the constraints defined from the external circumstances (number and kind of interactions, kind of categories, and so on).

In practice, both these elements seem to suggest how the role of categories becomes relevant when either decreases and degrades the knowledge within the analyzed system (before the interaction with the trustor) or is reduced the transferred knowledge (to the trustor).

Let us explain better. The *first effect* shows how the reliability of category's trustworthiness (that will be inherited by its members) depends, of course, from the variability of the behavior among the class members. There may be classes where all the members are very correct and competent, other classes where there is a very high variance: in this last case our betting on a member of that class is quite risky.

The *second effect* can be described with the fact that each agent, reducing the number of interactions with the other agents in the explorative phase, will have relevantly less information with respect to the individual agents. At the same time its knowledge with respect to categories does not undergo a significant decline given that categories' performances derive from several different agents.

The *third effect* can be explained with the fact that reducing the number of queried trustees, the trustor will receive with decreasing probability information about the more trustworthy individual agents in the domain, while information on categories, maintains a good level of stability also reducing the number of queried agents, thanks to greater robustness of these structures.

Resuming, the above pictures clearly show how, when the quantity of information (about the agents' trustworthiness exchanged in the system) decreases, it is better to rely on the categorial recommendations rather than individual recommendations.

This result reaches the point of highest criticality in the "all-in-one" case in which, as expected, the relevance of categories reach its maximal value.

V. CONCLUSION

Other works [9][2] show the advantages of using categorization to select trustworthy agents. In particular, how it were possible to attribute to a certain unknown agent, a value of trustworthiness with respect to a specific task, on the basis of its classification in, and membership to, one (/or more) category/ies. In practice, the role of generalized knowledge and prejudice (in the sense of pre-established judgment on the agents belonging to that category) has proven to determine the possibility to anticipate the value of unknown agents.

In this paper we have investigated the different roles that recommendations can play about individual agents and about categories of agents.

In this case the new agent introduced (called trustror) has a whole world of agents completely unknown to it, and ask for recommendations to a (variable) subset of agents for selecting an agent to whom delegate a task. The information received regards both individual agents and agents' categories. The informative power of these two kinds of recommendations depends on the previous interactions among the agents and also on the number agents queried by the trustor. However, there are cases in which information about categories is more useful that information towards individual agents. In some sense this result complements the results achieved in [9][2] because here we have a more strict match between information on individual agents and information about categories of agents: We are measuring the quantity of information, about individual agents and categories, for evaluating when is better using *direct information* rather than *generalized information* or, vice versa, when is better using the positive power of prejudice. Our results show how in certain cases becomes essential the use of categorial knowledge for selecting qualified partners.

In this work we have in fact considered a closed world, with a fixed set of agents. This choice was based on the fact that we were interested to evaluate the relationships between knowledge about individual and knowledge about categories, for calibrating their roles and reciprocal influences. In future works we have to consider how, starting from the analysis of this study, could change the role of knowledge about categories in a situation of open world. We have also to consider the cases in which the recommendations are not so transparent but influenced by specific goals of the agents.

ACKNOWLEDGMENTS

This work is partially supported both by the Project PRISMA (PiattafoRme cloud Interoperabili per SMArt-government; Cod. PON04a2 A) funded by the Italian Program for Research and Innovation (Programma Operativo Nazionale Ricerca e Competitività 2007-2013) and by the project CLARA—CLoud plAtform and smart underground imaging for natural Risk Assessment, funded by the Italian Ministry of Education, University and Research (MIUR-PON).

REFERENCES

[1] Adomavicius, G., Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering (TKDE) 17, 734–749, 2005

[2] Burnett, C., Norman, T., and Sycara, K. 2010. Bootstrapping trust evaluations through stereotypes. In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'10). 241248.

[3] C. Burnett, T. J. Norman, and K. Sycara. Stereotypical trust and bias in dynamic multiagent systems. ACM Transactions on Intelligent Systems and Technology (TIST), 4(2):26, 2013.

[4] Castelfranchi C., Falcone R., Trust Theory: A Socio-Cognitive and Computational Model, John Wiley and Sons, April 2010.

[5] Conte R., and Paolucci M., 2002, Reputation in artificial societies. Social beliefs for social order. Boston: Kluwer Academic Publishers.

[6] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti. Improving Recommendation Quality by Merging Collaborative Filtering and Social Relationships. In Proc. of the International Conference on Intelligent Systems Design and Applications (ISDA 2011) , Córdoba, Spain, IEEE Computer Society Press, 2011

[7] P De Meo, E Ferrara, D Rosaci, and G Sarné. Trust and Compactness of Social Network Groups. IEEE Transactions on Cybernetics, PP:99, 2014

[8] Falcone R, Castelfranchi C, Generalizing Trust: Inferencing Trustworthiness from Categories. In: TRUST 2008 - Trust in Agent Societies, 11th International Workshop, TRUST 2008. Revised Selected and Invited Papers (Estoril, Portugal, 12-13 May 2008). Proceedings, pp. 65 - 80. R. Falcone, S. K. Barber, J. Sabater-Mir, M. P. Singh (eds.). (Lecture Notes in Artificial Intelligence, vol. 5396). Springer, 2008.

[9] Falcone R., Piunti, M., Venanzi, M., Castelfranchi C., (2013), From Manifesta to Krypta: The Relevance of Categories for Trusting Others, in R. Falcone and M. Singh (Eds.) Trust in Multiagent Systems, ACM Transaction on Intelligent Systems and Technology, Volume 4 Issue 2, March 2013

[10] H. Fang, J. Zhang, M. Sensoy, and N. M. Thalmann. A generalized stereotypical trust model. In Proceedings of the 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pages 698–705, 2012.

[11] G. Guo, J. Zhang and N. Yorke-Smith, Leveraging Multiviews of Trust and Similarity to Enhance Clustering-based Recommender Systems, Knowledge-Based Systems, accepted, 2014

[12] Huynh, T.D., Jennings, N. R. and Shadbolt, N.R. An integrated trust and reputation model for open multi-agent systems. Journal of Autonomous Agents and Multi-Agent Systems, 13, (2), 119-154., 2006

[13] P. Lops, M. Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in Recommender Systems Handbook. Springer, pp. 73–105, 2011.

[14] P. Massa, P. Avesani, Trust-aware recommender systems, RecSys '07: Proceedings of the 2007 ACM conference on Recommender systems, 2007

[15] S. Ramchurn, N. Jennings, Carles Sierra, and Lluis Godo. Devising a trust model for multi-agent interactions using confidence and reputation. Applied Artificial Intelligence, 18(9-10):833-852, 2004.

[16] Sabater-Mir, J. 2003. Trust and reputation for agent societies. Ph.D. thesis, Universitat Autonoma de Barcelona.

[17] M. Sensoy, B. Yilmaz, and T. J. Norman. STAGE: Stereotypical trust assessment through graph extraction. Computational Intelligence, 2014.

[18] C. Than and S. Han, Improving Recommender Systems by Incorporating Similarity, Trust and Reputation, Journal of Internet Services and Information Security (JISIS), volume: 4, number: 1, pp. 64-76, 2014

[19] Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

[20] Yolum, P. and Singh, M. P. 2003. Emergent properties of referral systems. In Proceedings of the 2nd International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS'03).